

# Synthesising Graphical Theories

Aleks Kissinger

## Abstract

In recent years, diagrammatic languages have been shown to be a powerful and expressive tool for reasoning about physical, logical, and semantic processes represented as morphisms in a monoidal category. In particular, categorical quantum mechanics, or “Quantum Picturalism”, aims to turn concrete features of quantum theory into abstract structural properties, expressed in the form of diagrammatic identities. One way we search for these properties is to start with a concrete model (e.g. a set of linear maps or finite relations) and start composing generators into diagrams and looking for graphical identities.

Naïvely, we could automate this procedure by enumerating all diagrams up to a given size and check for equalities, but this is intractable in practice because it produces far too many equations. Luckily, many of these identities are not primitive, but rather derivable from simpler ones. In 2010, Johansson, Dixon, and Bundy developed a technique called *conjecture synthesis* for automatically generating conjectured term equations to feed into an inductive theorem prover. In this extended abstract, we adapt this technique to diagrammatic theories, expressed as graph rewrite systems, and demonstrate its application by synthesising a graphical theory for studying entangled quantum states.

## 1 Introduction

Monoidal categories can be thought of as theories of generalised (physical, logical, semantic, ...) processes. In particular, they provide an abstract setting for studying how processes (i.e. morphisms) behave when they are composed in time (via the categorical composition  $\circ$ ) and in space (via the monoidal product  $\otimes$ ). Categorical quantum mechanics, a program initiated by Abramsky and Coecke in 2004, takes this notion seriously by showing that many phenomena in quantum mechanics can be completely understood at the abstract level of monoidal categories carrying certain extra structure [1]. This notion has also had success in reasoning about stochastic networks [7], illustrative “toy theories” of physics [3], and even linguistics [6]. In most of these contexts, we make extensive use of spacial “swap” operations (symmetries) and temporal “feedback” operations (traces). Formally, this means we are working with a particular kind of monoidal category called a symmetric traced category.

Working with large compositions of morphisms in a symmetric traced category using traditional, term-based expressions quickly becomes unwieldy, as one attempts to describe an inherently two-dimensional structure using a (one-dimensional) term language. *Diagrammatic* languages thus provide a natural alternative.

String diagrams represent morphisms as “boxes” with various inputs and outputs, and compositions as “wires” connecting those boxes. They provide a way of visualising composition that is both highly intuitive and completely rigorous. They were originated by Penrose in the 1970s as a way to describe contractions of (abstract) tensors [18]. These abstract tensor systems closely resembled traced monoidal categories, so perhaps unsurprisingly, Joyal and Street were able to use string diagrams in 1991 to construct free symmetric traced categories [13]. In the context of categorical quantum mechanics, complementary observables [2], quantum measurements [9], and entanglement [4] can be studied using string diagrams and diagrammatic identities.

Joyal and Street’s formalisation of string diagrams uses topological graphs with some extra structure. While this method comes very close to the intuitive notion of a string diagram, it is not obvious how one could automate the creation and manipulation of these topological objects

in a computer program. For that reason, Dixon, Duncan, and Kissinger developed a discrete representation of string diagrams called *string graphs*. Unlike their continuous counterparts, it is straightforward to automate the manipulation of string graphs using graph rewrite systems. Using a graph rewrite system, we can draw conclusions about any concrete model of that system. Since the amount of data needed to represent a morphism concretely often grows exponentially in the number of inputs and outputs, we can gain potentially huge computational benefits by manipulating these morphisms using rewrite rules instead of concrete calculation.

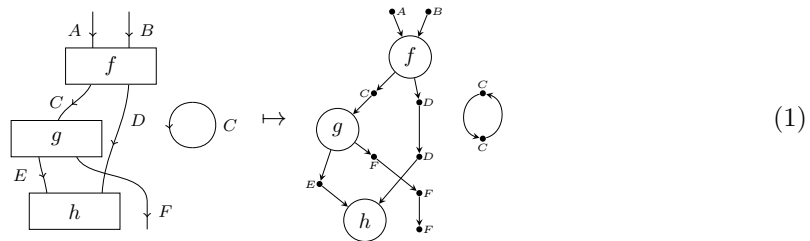
We might also ask if this can be done in reverse. Consider a situation where we have a collection of generators as well as concrete realisations of those generators (e.g. as linear maps, finite relations, etc.) and we wish to discover the algebraic structure of these generators. One way to find new rules is to start enumerating all string graphs involving those generators, evaluating them concretely, and checking for identities. This process becomes intractable very quickly, as the number of distinct string graphs over a set of generators grows exponentially with size. However, in 2010, Johansson, Dixon, and Bundy provided a clever way of avoiding enumerating redundant rewrite rule candidates for a term rewrite theory, which they called *conjecture synthesis* [12]. This technique keeps a running collection of rewrite rules and only enumerates terms that are *irreducible* with respect to this collection of rules. This simple condition is surprisingly effective at curbing the exponential blow-up of the rewrite system synthesis procedure, and provides the basis for a piece of software developed by the authors called IsaCoSy [11] (for ISAbelle COncecture SYNthesis).

The main contribution of this paper is an algorithm that generates new graphical theories using an adapted conjecture synthesis procedure. This adapted procedure has been implemented in a tool called QuantoCoSy, built on string graph rewriting platform called Quantomatic [15]. We provide an example of applying QuantoCoSy to generate a graphical theory used in the study of quantum entanglement. We use as a basis the string graph (aka open-graph) formalism developed in [8].

After briefly reviewing string graphs and string graph rewriting in section 2, we show in section 3 how, given a valuation of the generators of a string graph, the graph itself can be evaluated as a linear map using tensor contraction. The graphical conjecture synthesis procedure is then described in section 4. In section 5, we look at a test case where our theory synthesis software, QuantoCosy, is used to synthesise a simple graphical theory of entangled states.

## 2 Background: String Graphs

In this section, we summarise the construction of the category  $\mathbf{SGraph}_T$  of string graphs parametrised by a monoidal signature. Details of how this can be done formally in the context of adhesive categories can be found in [8]. Our goal in defining string graphs is to represent diagrammatic theories using graph rewrite systems. To do this, we need to re-cast (topological) string diagrams as typed graphs. We do this by replacing *wires* with chains of special vertices called *wire-vertices*. We also introduce *node-vertices*, which represent “logical” nodes (aka boxes) in the diagram.



A (small, strict) monoidal signature  $T$  defines a set of generators that can be interpreted in a monoidal category. Let  $w(X)$  be the set of lists over a set  $X$ .  $T$  consists of a set  $\mathcal{M}$  of *morphisms*, a set  $\mathcal{O}$  of objects and functions  $\text{dom}, \text{cod} : \mathcal{M} \rightarrow w(\mathcal{O})$  that assign input and output types for each morphism. Note in particular that  $T$  says nothing about composition.

**Definition 2.1.** The category  $\mathbf{SGraph}_T$  has as objects string graphs with node-vertex and wire-vertex types given by  $T$ , and as arrows graph homomorphisms respecting these types.

From hence forth, we will refer to chains of wire vertices simply as wires, when there is no ambiguity. There are a few things to note here. Firstly, wire-vertices carry the (object) type of a wire, so every connection between two node-vertices must contain at least one node-vertex. We do not allow wires to split or merge, so wire-vertices must have at most one input and one output. When our generators are non-commutative, we distinguish inputs and outputs using edge types  $(\text{in}_1, \text{in}_2, \text{out}_1, \text{out}_2, \dots)$ .

This category can be defined as a full subcategory of the slice category  $\mathbf{Graph}/G_T$ , where  $G_T$  is called the *derived typegraph* of a monoidal signature  $T$ . Furthermore,  $\mathbf{SGraph}_T$  is a *partial adhesive category*, as defined in [8]. Such a category provides the mechanisms we need to perform graph matching and rewriting in the following sections. For details, see [8].

**Definitions 2.2.** For a string graph  $G$ , let  $\omega(G)$  be the set of wire-vertices and  $\eta(G)$  the set of node-vertices. If a wire-vertex has no in-edges, it is called an *input*. We write the set of inputs of a string graph  $G$  as  $\text{In}(G)$ . Similarly, a wire-vertex with no out-edges is called an *output*, and the set of outputs is written  $\text{Out}(G)$ . The inputs and outputs define a string graph's *boundary*, written  $\text{Bound}(G)$ . If a boundary vertex has no in-edges and no out-edges, (it is both an input and output) it is called an *isolated wire-vertex*.

## 2.1 Plugging and Rewriting for String Graphs

There are two important, related operations we wish to apply to string graphs. The first is plugging, where input and output wire-vertices are glued together, and the second is string graph rewriting, where parts of the string graph are cut out and replaced with other graphs.

**Definition 2.3.** For a string graph  $G$  and wire-vertices  $x \in \text{In}(G)$  and  $y \in \text{Out}(G)$ , let  $G/(x,y)$  be a new string graph with the wire-vertices  $x$  and  $y$  identified. This is called the *plugging* of  $(x,y)$  in  $G$ , and the new vertex  $(x \sim y) \in G/(x,y)$  is called the *plugged wire-vertex*.

We can compose string graphs by performing pluggings on the disjoint union:  $(G + H)/(x,y)$  for  $x \in \text{Out}(G)$ ,  $y \in \text{In}(H)$ . We can plug many wires together by performing many pluggings one after another. Equivalently, we can perform a pushout along a discrete graph, respecting certain conditions. When there can be no confusion, we also call this a plugging.

**Definition 2.4.** A pushout of a span  $G \xleftarrow{p} K \xrightarrow{q} H$  is called a *plugging* if  $K$  is a disconnected graph of wire-vertices,  $p$  and  $q$  are mono, and for all  $k \in K$ ,  $p(k) \in \text{Bound}(G)$ ,  $q(k) \in \text{Bound}(H)$  and  $p(k) \in \text{In}(G) \Leftrightarrow q(k) \in \text{Out}(H)$ .

We perform rewrites on string graphs using the *double pushout* (DPO) graph rewriting technique. We first define a string graph rewrite rule as a pair of string graphs sharing a boundary. It will be convenient to formalise this as a particular kind of span.

**Definition 2.5.** A *string graph rewrite rule*  $L \rightarrow R$  is a span of monomorphisms  $L \xleftarrow{b_1} B \xrightarrow{b_2} R$  such that  $b_1(B) = \text{Bound}(L)$ ,  $b_2(B) = \text{Bound}(R)$  and for all  $b \in B$ ,  $b_1(b) \in \text{In}(L) \Leftrightarrow b_2(b) \in \text{In}(R)$ .

To perform rewriting, we first introduce a notion of matchings. These are special monomorphisms that only connect to the rest of the graph via their boundary.

**Definition 2.6.** A *matching* of a string graph  $L$  on another string graph  $G$  is a monomorphism  $m : L \rightarrow G$  such that whenever an edge not in the image of  $m$  is adjacent to  $m(v)$  for some vertex  $v$  in  $L$ ,  $v$  must be in the boundary  $\text{Bound}(L)$  of  $L$ . When such a matching exists, we say  $L$  (or any rule with  $L$  as its LHS) *matches*  $G$ .

Note that, for the class of rewrite rules considered (i.e. those of the form of Definition 2.5), this corresponds to the notion of matching defined by Ehrig et al [19]. Once a matching is found, the occurrence of  $L$  in  $G$  can be replaced with  $R$ . However, some care must be taken to ensure that  $R$  is reconnected to the remainder of  $G$  in the same location as  $L$  was previously. We begin by removing the *interior* of  $L$  in  $G$ , i.e. the part of  $L$  that is not in the boundary. We do this by finding a *context* graph  $G'$  such that when  $L$  and  $G'$  are plugged together along the boundary of  $L$ , the total graph  $G$  is obtained. This is known as computing the *pushout complement* of  $B \xrightarrow{b_1} L \xrightarrow{m} G$ . Since  $G$  is the result of gluing  $L$  and  $G'$  together along  $B$ , we can replace  $L$  with  $R$  by performing a second plugging, this time of  $R$  and  $G'$  along  $B$ .

$$\begin{array}{ccccc}
 L & \xleftarrow{\quad} & B & \xrightarrow{\quad} & R \\
 m \downarrow & \lrcorner & \downarrow m' & \lrcorner & \downarrow \\
 G & \xleftarrow{\quad} & G' & \xrightarrow{\quad} & H
 \end{array} \tag{2}$$

In other words, we express  $G$  as a plugging of two graphs  $L$  and  $G'$ , then compute  $H$  by plugging  $R$  into  $G'$ . This completes the rewrite of  $G$  into  $H$ . Note that even in a category with all pushouts, pushout complements need not exist or be unique. However, the following theorem, proved in [8], shows that DPO rewriting is always well-defined for string graph matchings and rewrite rules.

**Theorem 2.7.** Let  $L \xleftarrow{b_1} B \xrightarrow{b_2} R$  be a boundary span and  $m : L \rightarrow G$  a matching. Then  $B \xrightarrow{b_1} L \xrightarrow{m} G$  has a unique pushout complement  $G'$  and both of the pushout squares in diagram (2) exist and are preserved by the embedding of  $\mathbf{SGraph}_T$  into  $\mathbf{Graph}/G_T$ .

We will also find the following lemma useful when combining matching and string graph enumeration. It shows that performing a plugging only creates new matchings *local* to the plugged vertex.

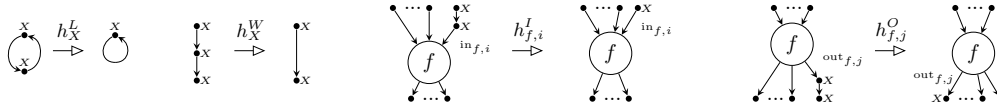
**Lemma 2.8.** Let  $x \in \text{In}(G)$ ,  $y \in \text{Out}(G)$ , then let  $m : L \rightarrow G/(x,y)$  be a matching such that the plugged vertex  $x \sim y$  is not in the image of  $m$ . Then there exists a matching  $m' : L \rightarrow G$ .

*Proof.* Let  $q : G \rightarrow G/(x,y)$  be the quotient map. Then, pulling back  $q$  over  $m$  is just the restriction of  $q$  to the image of  $m$ . Since the image of  $m$  does not contain the plugged vertex,  $q$  restricts to an isomorphism  $r$ . It is straightforward to show that  $m' \circ r^{-1}$  is a matching.  $\square$

**Definition 2.9.** A set of string graph rewrite rules is called a *string graph rewrite system*. For a string graph rewrite system  $\mathbb{S}$ , we write  $G \rightarrow_{\mathbb{S}} H$  if there exists a rule  $L \rightarrow R \in \mathbb{S}$  and a matching  $m : L \rightarrow G$  such that  $G$  can be rewritten to  $H$  using the DPO diagram given by diagram (2). Let  $\xrightarrow{*}_{\mathbb{S}}$  be the reflexive, transitive closure of  $\rightarrow_{\mathbb{S}}$  and  $\xleftrightarrow{*}_{\mathbb{S}}$  be the associated equivalence relation.

The vigilant reader will notice that we allow any number of wire-vertices to occur within a wire (all with the same type). However, the number of wire-vertices makes no semantic difference in the interpretation of a string graph. Therefore, we will consider two string graphs to be semantically equivalent if they only differ in the length of their wires. This equivalence relation is called *wire-homeomorphism*. We can formalise wire-homeomorphism as a string graph rewrite system.

**Definition 2.10.** For a monoidal signature  $T = (\mathcal{O}, \mathcal{M}, \text{dom}, \text{cod})$ , the rewrite system  $\mathbb{H}$  is defined as follows. For every  $X \in \mathcal{O}$ , we define a loop contraction rule  $h_X^L$  and a wire contraction rule  $h_X^W$ . For every  $f \in \mathcal{M}$  and  $0 \leq i < \text{Length}(\text{dom}(f))$ ,  $0 \leq j < \text{Length}(\text{cod}(f))$ , we define an input contraction rule  $h_{f,i}^I$  and an output contraction rule  $h_{f,j}^O$ .



It was shown in [8] that this is a confluent, terminating rewrite system. Formal forms are called *reduced string graphs*, and contain only one wire-vertex on every wire.

### 3 Interpreting String Graphs as Tensor Contractions

Describing the contractions of many tensors was the main reason Penrose introduced string diagram notation, so perhaps unsurprisingly, there is a natural way to interpret a string graph as a contraction of tensors. We first give the monoidal signature  $T = (\mathcal{O}, \mathcal{M}, \text{dom}, \text{cod})$  a *valuation*  $v : T \rightarrow \mathbf{Vect}_{\mathbb{C}}$ . This is a monoidal signature homomorphism from  $T$  to  $\mathbf{Vect}_{\mathbb{C}}$  that assigns to each of the objects  $o \in \mathcal{O}$  a vector space and each of the morphisms  $f \in \mathcal{M}$  a linear map  $v(f)$ . We can then regard the linear map  $v(f)$  as a tensor with a lower index for every input to  $f$  in  $T$  and an upper index for every output.

A string graph can then be interpreted as a big tensor contraction by interpreting wire-vertices as identities (i.e. the Dirac delta tensor  $\delta_i^j$ ), node-vertices as linear maps  $v(f)$ , for  $f$  the type of the vertex, and edges as sums over an index.

The equation shows the evaluation of a string graph as a tensor contraction. On the left, a string graph with two nodes  $f$  and  $g$  is shown. Node  $f$  has inputs  $k, l$  and output  $o$ . Node  $g$  has inputs  $m, n$  and output  $p$ . A wire connects  $o$  to  $p$ . On the right, the corresponding tensor contraction is shown, with indices grouped into inputs, outputs, nodes, and wires.

$$\text{String Graph} \mapsto \underbrace{\delta_{i_1}^k \delta_{i_2}^l}_{\text{inputs}} \underbrace{\delta_m^{j_1} \delta_n^{j_2} \delta_o^{j_3}}_{\text{outputs}} \underbrace{[v(f)]_{k,l}^{p,o} [v(g)]_q^{m,n}}_{\text{nodes}} \underbrace{\delta_p^q \delta_r^r}_{\text{wires}}$$

We are using the Einstein summation convention (repeated indices are summed over), and we have labeled edges in the string graph with their corresponding indices. Like wire-vertices themselves, the  $\delta$  maps are used mainly just for book-keeping. They maintain the correct order of inputs and outputs, define circles, and connect more interesting tensors together.

This gives the evaluation of a string diagram  $G$  as a linear map  $\llbracket G \rrbracket$  in  $\mathbf{Vect}_{\mathbb{C}}$ . This is a special case of a much more general construction. In [14], Kissinger showed that string graphs could be used to form the free symmetric traced category over a monoidal signature. As a consequence, for *any* symmetric traced category  $\mathcal{C}$ , a monoidal signature homomorphism  $v : T \rightarrow \mathcal{C}$  lifts uniquely to an evaluation functor from the category of string graphs to  $\mathcal{C}$ .

### 4 Conjecture Synthesis for String Graph Theories

We will now describe a synthesis procedure for graphical identities in the spirit of IsaCoSy [11]. A notable difference is that, rather than passing conjectured identities off to an inductive

theorem prover, we simply evaluate them using a given valuation of the generators of the theory.

The key to this procedure is that it maintains a set of reduction rules  $\mathbb{S}$  throughout, and avoids enumerating string graphs that are reducible with respect to  $\mathbb{S}$ . We refer to such graphs as *redexes*.

The theory synthesis procedure takes as input a string-graph signature  $T$  along with an  $(m, n)$ -tensor for every box in  $T$  with  $m$  inputs and  $n$  outputs. It also takes an initial set of rewrite rules  $\mathbb{S}$  and a reduction ordering  $\kappa$ . This is a function from string graphs to  $\mathbb{N}$  where  $G \cong H \Rightarrow \kappa(G) = \kappa(H)$  and  $G \rightarrow_{\mathbb{S}} H \Rightarrow \kappa(G) > \kappa(H)$ . We shall also chose  $\kappa$  to be non-increasing in the number of node-vertices and wire-vertices in a string graph. We will also maintain a set  $\mathbb{K}$  (initially empty) of *congruences*, but these cannot be used for redex-elimination.

In the term case, a single round of synthesis is parametrised by two natural numbers: the maximum term size (or term depth) and the maximum number of free variables occurring in the term. For string graphs, we parametrise a run with four natural numbers: the number of inputs  $M$ , the number of outputs  $N$ , the maximum number of pluggings  $P$ , and the maximum number of node-vertices  $Q$ . We shall refer to string graphs with  $M$  inputs,  $N$  outputs, and up to  $P, Q$  pluggings and node-vertices as *string graphs of size*  $(M, N, P, Q)$ .

First, we define a string graph enumeration procedure that avoids reducible expressions with respect to our initial rewrite system  $\mathbb{S}$ . Suppose we have a string graph signature  $T = (\mathcal{O}, \mathcal{M}, \text{dom}, \text{cod})$ , then we can define a string graph for every morphism in  $T$ , called its generator. For each  $f \in \mathcal{M}$ , this is the smallest string graph containing a node-vertex of type  $f$ . For completeness, we also include the identity generator, which is just two connected wire-vertices. For an example of a set of generators, see (3) in the next section.

We enumerate string graphs by starting with disconnected string graphs, i.e. disjoint unions of generators, and performing pluggings to connect separate components together.

**Definition 4.1.** For a string graph  $G$ , let  $\mathbf{p}(G)$  be the set  $\text{In}(G) \times \text{Out}(G)$ . A pair  $(x, y) \in \mathbf{p}(G)$  defines a particular plugging  $G/(x, y)$ . Two pluggings in  $(x, y), (x', y') \in \mathbf{p}(G)$  are called *similar* if there exists an isomorphism  $\phi : G/(x, y) \cong G/(x', y')$  that is identity on node-vertices. Let  $\mathbf{p}_{(x, y)}(G) \subseteq \mathbf{p}(G)$  be the set of all pairs  $(x', y')$  similar to  $(x, y)$ .

Let  $\mathcal{D}(M, N, Q)$  be the set of all disconnected string graphs with  $M$  inputs,  $N$  outputs and up to  $Q$  node-vertices. The procedure ENUM takes as input a graph, a set of (distinguishable) pluggings and the number of pluggings left to do. Note that each plugging decreases  $|\text{In}(G)|$  and  $|\text{Out}(G)|$  by 1. So, if we start with  $G \in \mathcal{D}(M + p, N + p, Q)$  and before  $p$  pluggings, we will get a string graph with  $M$  inputs and  $N$  outputs.

```

1: procedure ENUM( $\Pi, G, p$ )
2:   if  $p = 0$  then
3:     save  $G$ 
4:     exit
5:   else if  $\Pi = \{\}$  then
6:     exit
7:   end if
8:   let  $(x, y) \in \Pi$ 
9:   if no rule in  $\mathbb{S}$  matches  $G/(x, y)$ , local to  $(x, y)$  then
10:     ENUM( $\Pi - \{(x, y)\}, G/(x, y), p - 1$ )
11:   end if
12:   let  $\Pi' = \Pi - \mathbf{p}_{(x, y)}(G)$ 

```

```

13:   ENUM( $\Pi'$ ,  $G$ ,  $p$ )
14: end procedure
15:
16: for all  $0 \leq p \leq P$ ,  $G \in \mathcal{D}(M + p, N + p, Q)$  do
17:   ENUM( $\mathfrak{p}(G)$ ,  $G$ ,  $p$ )
18: end for

```

The procedure ENUM recurses down two branches, one where we decide to do a particular plugging  $(x, y) \in \Pi$  (line 10), and one where we decide not to (line 13). Line 12 prevents us from deciding not to do a certain plugging then later deciding to do one that is similar, as this will enumerate redundant graphs.

The crucial redex-elimination step in the algorithm occurs at line 9. Recall that a rule  $L \rightarrow R$  matches  $G/\!(x, y)$  local to  $(x, y)$  if the matching  $m : L \rightarrow G/\!(x, y)$  contains the plugged vertex  $(x \sim y) \in G/\!(x, y)$  in its image. Assuming each of our generators is irreducible, it suffices to consider only matches of rules local to the most recent plugging in order to eliminate all reducible expressions. By Lemma 2.8, if a rule has a matching on  $G/\!(x, y)$  that is not local to  $(x, y)$ , it already has a matching on  $G$ . Therefore, the condition on line 9 guarantees no redexes, with respect to  $\mathbb{S}$ , will be enumerated.

Once all of the irreducible graphs of size  $(M, N, P, Q)$  have been enumerated, we update the rewrite system as follows. First, we evaluate the string graphs as linear maps and organise them into equivalence classes, up to scalar factors and permutations of inputs and outputs. We then filter out any remaining isomorphic graphs in each equivalence class and identify the set of minimal string graphs  $C' \subseteq C$  with respect to  $\kappa$ . Choose a representative  $G_0$  of  $C'$ . Finally, we add new reductions  $G \rightarrow G_0$  to  $\mathbb{S}$  for all  $G \in C - C'$ . Add congruences  $G' \rightarrow G_0$  and  $G_0 \rightarrow G'$  for all  $G' \in C' - \{G_0\}$  to  $\mathbb{K}$ .

We postpone filtering out isomorphic string graphs until after enumeration because tensor contraction is fast, and two string graphs will not be isomorphic unless they are in the same equivalence class. In order to get a well-behaved rewrite system, we should choose  $\kappa$  so that there are very few congruences, if any. Obviously, if we choose  $\kappa$  to be strict (for all  $G \not\cong H$ ,  $\kappa(G) < \kappa(H)$  or  $\kappa(G) > \kappa(H)$ ), there will be no congruences. While strict reduction orderings for graphs are much more difficult to compute than their term analogues, this may be tractable if we adapted our graph enumeration procedure to only produce canonical representatives of isomorphism classes of string graphs (cf. [10, 16]).

Once a single run of the synthesis procedure is complete, we can re-run the procedure using larger values of  $M$ ,  $N$ ,  $P$ , and  $Q$  as well as the updated rewrite system  $\mathbb{S}$ . Using this growing collection of reductions can be very effective in stemming the exponential blow-up in both the number of string graphs that need to be enumerated and the number of rewrite rules found.

**Theorem 4.2.** *Applying the synthesis procedure for a series of runs given by  $\{(M_i, N_i, P_i, Q_i)\}$  yields a rewrite system  $\mathbb{S} \cup \mathbb{K}$  that is complete for all graphs of size  $(M_i, N_i, P_i, Q_i)$  for some  $i$ . Furthermore, if  $\mathbb{K} = \{\}$  the rewrite system yields unique normal forms for graphs of the given sizes.*

*Proof.* For completeness, suppose there are string graphs  $G$  and  $H$  of size  $(M_i, N_i, P_i, Q_i)$  such that  $\llbracket G \rrbracket = \llbracket H \rrbracket$ . Since  $\mathbb{S}$  is a terminating rewrite system, we can normalise  $G$  and  $H$  with respect to  $\mathbb{S}$ . Since  $\kappa$  is non-increasing on node-vertices and wire-vertices, the associated normal forms  $G'$  and  $H'$  are also of size  $(M_i, N_i, P_i, Q_i)$ , so both graphs will have been enumerated. Therefore, either  $G' \cong H'$  or  $(G' \rightarrow H') \in \mathbb{K}$ . For unique normal forms, note that when  $\mathbb{K}$  is empty, only  $G' \cong H'$  is possible.  $\square$

Of course, we are interested in graphs of all sizes. However, completeness is not expected (and sometimes not desirable) in these cases, as the model or models used to synthesise the theory may be degenerate in some sense as examples of some algebraic structure. To ensure unique normal forms, we would need to ensure confluence of  $\mathbb{S}$ . Unlike termination, this does not come for free, as there may exist critical pairs that are larger than any of the graphs synthesised. However, the synthesis procedure could be used in conjunction with a graphical variant of Knuth-Bendix completion to increase the chances of obtaining a confluent rewrite system for graphs of all sizes.

## 5 Application and Results

The procedure described in section 4 has been implemented in a tool called *QuantoCosy*. This sits on top a general framework for string graph rewriting called *Quantomatic* [15]. We show in this section how *QuantoCosy* performs in synthesising an example theory, called the GHZ/W-calculus. For a detailed description of this theory and how it can be applied, see [5].

In the GHZ/W-calculus, we can express large, many-body entangled quantum states as compositions of simpler components. The key here is to use three-body states as algebraic “building blocks” for more complex states. A three-system, or tripartite entangled state, can be represented as a vector in  $H \otimes H \otimes H$  for some complex vector space  $H$ . For finite dimensions, we can equivalently express this vector as a linear map  $H \otimes H \rightarrow H$ . When  $H \cong \mathbb{C}^2$ , there are only two “interesting” tripartite entangled states, the GHZ state and the W state. Regarding these as linear maps from  $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^4$  to  $\mathbb{C}^2$ , they can be written as follows:

$$\begin{array}{c} \bullet \quad \bullet \\ \swarrow \quad \searrow \\ \circ \\ \downarrow \\ \bullet \end{array} := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \begin{array}{c} \bullet \quad \bullet \\ \swarrow \quad \searrow \\ \bullet \\ \downarrow \\ \bullet \end{array} := \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

These maps exhibit particularly nice algebraic identities. They can both be extended to *commutative Frobenius algebras*, which are unital, associative algebras over a vector space (or an object in some monoidal category) that have a strong *self-duality* property. In particular, they can be naturally associated with a co-unital, co-algebra that interacts well with the algebra. For pairs of Frobenius algebras, we can also always construct a special map called the *dualiser*. See [5] for details on how these things are defined algebraically.

The generators of the GHZ/W-calculus are: (i) the generators of both Frobenius algebras, (ii) the dualiser of the two algebras, and (iii) two zero vectors.

$$\left\{ \begin{array}{c} \bullet \\ \downarrow \\ \textcircled{\bullet} \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \swarrow \searrow \\ \textcircled{\bullet} \\ \swarrow \searrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \downarrow \\ \textcircled{\bullet} \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \downarrow \\ \textcircled{\bullet} \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \downarrow \\ \textcircled{\bullet} \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \downarrow \\ \blacksquare \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \triangle \\ \downarrow \\ 0 \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \downarrow \\ \nabla \\ \downarrow \\ \bullet \end{array}, \begin{array}{c} \bullet \\ \downarrow \\ \bullet \end{array} \right\} \quad (3)$$

As was the case for terms, filtering out redexes has a huge impact on the number of string graphs that need to be checked. In ten synthesis runs, we generate all of the GHZ/W rewrite rules with total node-vertices, pluggings, and inputs+outputs  $\leq 3$ . Using a naïve synthesis, this yielded 13,302 rewrite rules. Using the procedure described in the previous section, this yielded 132 rules, including all of the rules used to define the GHZ/W-calculus in [5]. A plot of the number of rewrite rules generated using a naïve graph enumeration algorithm against the number generated using the redex-elimination procedure across all 10 runs is provided in Figure 1.

## 6 Conclusion and Outlook

In this paper, we showed how string diagrams could be discretised into string graphs, which can be manipulated automatically using DPO graph rewriting. We then demonstrated how



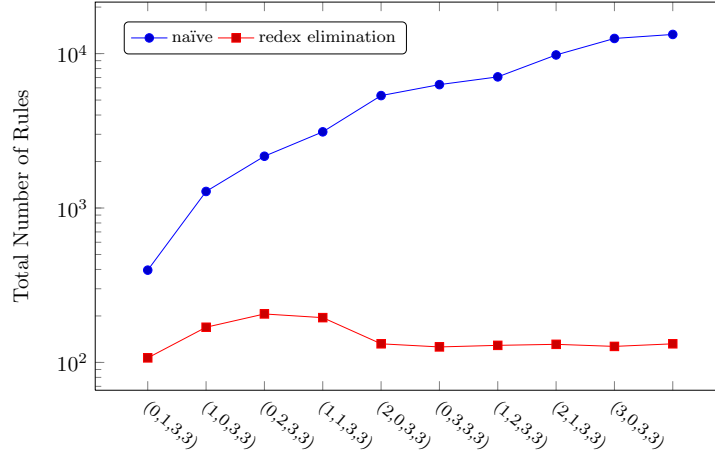


Figure 1: Number of rules synthesised across 10 runs of the synthesis procedure.

a technique developed for term theories called conjecture synthesis can be adapted to work for string graphs. Finally, we showed an example of the application of this technique to a real graphical theory, called the GHZ/W-calculus. The results there were promising, as we demonstrated an exponential drop-off in the number of extraneous rules generated when using redex-eliminating routine as compared to a naïve synthesis routine.

In the end, some of the 132 rules produced were seemingly arbitrary, whereas others reflect a real algebraic relationship between the GHZ-structure and the W-structure. There are various reasons, including simple aesthetics, why a human mathematician would take some of these rewrites to be valuable or interesting. QuantoCosy is completely ignorant to such considerations, as it only knows what rules to throw away, rather than which ones to highlight. However, this is already useful, as it provides a researcher with a pool of hundreds of rules to sort through, rather than tens of thousands.

There are various ways in which this can be improved further. One is to make the synthesis smarter by employing heuristics that search for particular classes of rewrites that are common to many algebraic structures. This has been done for terms using *scheme-based* conjecture synthesis [17], where the search procedure starts by attempting to prove familiar sets of identities (associativity and commutativity of a binary operation, for instance) before moving to less familiar conjectures.

Another way to improve the quality and conciseness of the synthesis output would be to refine the synthesised rules using Knuth-Bendix completion. This could prove a powerful method for automatically producing new rewrite rules for *pattern graphs* [14]. Pattern graphs are used to describe infinite families of string graphs that have some repeated substructure. Starting with a rewrite system that contains some pattern graph rewrite rules that are previously known, and some concrete rules that are produced using conjecture synthesis, one can obtain new pattern graph rewrites by performing critical pair analysis and completion.

Finally, and most importantly, this procedure can be improved and its scope broadened by looking at different kinds of concrete models. As mentioned briefly in section 3, string graphs can be used to construct the free symmetric traced category over a signature. So, any interpretation of a signature in a concrete symmetric traced category can be lifted uniquely to an evaluation of string graphs. Examples of such categories where one might wish to form

models are:  $\mathbf{Mat}(F)$  of matrices over a finite field,  $(\mathbf{FRel}, \times)$  of sets and finite relations with the “wave” style tensor and trace,  $(\mathbf{FRel}, +)$  of sets and finite relations with the “token” style tensor and trace, and products of concrete symmetric traced categories (i.e. giving valuations in many models simultaneously).

The support of a theory synthesis tool allows a mathematician to very quickly get a picture of a theory by specifying a few generators, and opens to the way for experimentation and rapid development of a wide variety of new theories.

## References

- [1] S. Abramsky and B. Coecke. A categorical semantics of quantum protocols. In *Proceedings from LiCS*, arXiv:quant-ph/0402130v5, 2004.
- [2] B. Coecke and R. Duncan. Interacting quantum observables. In *Proceedings from ICALP*, pages 298–310, 2008.
- [3] B. Coecke and B. Edwards. Spekkens’s toy theory as a category of processes. arXiv:1108.1978v1 [quant-ph], 2011.
- [4] B. Coecke and A. Kissinger. The compositional structure of multipartite quantum entanglement. In *Automata, Languages and Programming*, volume 6199 of *Lecture Notes in Computer Science*, pages 297–308. Springer, 2010.
- [5] B. Coecke and A. Kissinger. The compositional structure of multipartite quantum entanglement. arXiv:1002.2540v2 [quant-ph], 2010.
- [6] B. Coecke, M. Sadrzadeh, and S. Clark. Mathematical foundations for a compositional distributional model of meaning. arXiv:1003.4394v1 [cs.CL], 2010.
- [7] B. Coecke and R. W. Spekkens. Picturing classical and quantum bayesian inference. arXiv:1102.2368v1 [quant-ph], 2011.
- [8] L. Dixon and A. Kissinger. Open Graphs and Monoidal Theories. arXiv:1007.3794v1 [cs.LO], 2010.
- [9] R. Duncan and S. Perdrix. Rewriting measurement-based quantum computations with generalised flow. In *Proceedings of the 37th international colloquium conference on Automata, languages and programming: Part II*, ICALP’10, pages 285–296, Berlin, Heidelberg, 2010. Springer-Verlag.
- [10] L. Goldberg. Efficient algorithms for listing unlabeled graphs. *Journal of Algorithms*, 13(1):128–143, 1992.
- [11] M. Johansson, L. Dixon, and A. Bundy. IsaCoSy. <http://dream.inf.ed.ac.uk/projects/isaplanner/wiki/IsaCoSy>.
- [12] M. Johansson, L. Dixon, and A. Bundy. Conjecture Synthesis for Inductive Theories. *Journal of Automated Reasoning*, 2010.
- [13] A. Joyal and R. Street. The geometry of tensor calculus I. *Advances in Mathematics*, 88:55–113, 1991.
- [14] A. Kissinger. *Pictures of Processes: Automated Graph Rewriting for Monoidal Categories and Applications to Quantum Computing*. PhD thesis, University of Oxford (awaiting review), 2012.
- [15] A. Kissinger, A. Merry, L. Dixon, R. Duncan, M. Soloviev, and B. Frot. Quantomatic. <https://sites.google.com/site/quantomatic/>, 2011.
- [16] B. D. McKay. Isomorph-free exhaustive generation. *J Algorithms*, 26:306–324, 1998.
- [17] O. Montano-Rivas, R. McCasland, L. Dixon, and A. Bundy. Scheme-based synthesis of inductive theories. *Advances in Artificial Intelligence*, pages 348–361, 2010.
- [18] R. Penrose. Applications of negative dimensional tensors. In *Combinatorial Mathematics and its Applications*, pages 221–244. Academic Press, 1971.
- [19] U. Prange, H. Ehrig, and L. Lambers. Construction and properties of adhesive and weak adhesive high-level replacement categories. *Applications of Categorical Structures*, 16:365–388, 2008.

This figure "quanto\_cosy\_screen.png" is available in "png" format from:

<http://arxiv.org/ps/1202.6079v2>